DOCUMENT RESUME

ED 152 692 SP 012 228

AUTHOR Maguire, Thomas O.; Haig, Brian D.
TITLE Problems of Control in Non-Experimental Educational Research. Research and Information Report 75-1.
INSTITUTION Alberta Univ., Edmonton. Div. of Educational Research Services.
PUB DATE May 75
NOTE 22p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Conceptual Schemes; Data Analysis; *Educational Research; *Fundamental Concepts; *Research Methodology; *Statistical Analysis
IDENTIFIERS *Nonexperimental Research; *Nuisance Variables; Research as a Field of Study

ABSTRACT

An expansion is made of P. E. Meehl's observations on nuisance variables and the ex post facto research design. Implications these observations have for educational research design are discussed. Meehl (1970) considered that suggested resolutions of problems in "not truly experimental" (NTE) research (i.e., matching, partial correlation, and analysis of covariance) may generate more problems than they resolve. Three reasons he notes are systematic unmatching, unrepresentative subpopulations, and the causal arrow ambiguity. It is the first of these reasons to which the author devotes most of his attention. An extended statistical analysis of a hypothetical case involving achievement need leads to the conclusions, supported by L. J. Cronbach and Meehl (1955), that a construct derives its meaning from the network of relationships that connect it with other constructs and with observables. When nuisance variables are partialled out or controlled, the relationships among residual variables are not the same as the relationships among the original variables, and their construct status cannot be assumed to be the same. The author suggests that the most useful way to deal with the problem of interpreting NTE research is to live with nuisance variables by incorporating them into models of behavior, arguing that, if naturally occurring characteristics confound each other, it makes no sense to try to isolate individual effects. Rather, more observations across time should be made in an attempt to capture the interweaving of variables. (MJB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

research and information report

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Thomas O. Maguire

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to: SP | IR

In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

division of educational research
faculty of education
the university of alberta
edmonton, canada

Division of Educational Research Services
University of Alberta

RIR-75-1

PROBLEMS OF CONTROL IN NON-EXPERIMENTAL EDUCATIONAL RESEARCH

Thomas O. Maguire
University of Alberta
Edmonton, Alberta, Canada

Brian D. Haig
Canterbury University
Christchurch, New Zealand

May, 1975

PROBLEMS OF CONTROL IN NON-EXPERIMENTAL EDUCATIONAL RESEARCH

Each year throughout North America, hundreds of graduate students in education are exposed to the mysteries of the design and analysis of educational experiments. Using the now classic treatise of Campbell and Stanley (1963) they learn that many of the sources of invalidity that cloud the investigation of relationships among educational variables, can be controlled through the use of "True Experiments". They learn that in the basic true experiment, subjects are randomly assigned to groups, the groups are treated differentially, and observations are made to determine the effects of the treatment. The magic potion that controls the nuisance variables is random assignment.

In spite of the validity of the claims made for the use of true experiments in educational research, by far, the bulk of educational research is not truly experimental. Campbell and Stanley's pre and quasi experiments, the ex post facto experiments of Chapin (1955), case studies, field studies, and clinical methods are common (but not mutually exclusive) examples of not truly experimental (NTE) research methods.

There are many reasons why NTE research flourishes. Often people cannot be randomly assigned to groups for ethical reasons, as for example in looking at differences in visual acuity between deaf and hearing children. There may be administrative reasons that make random assignment impossible as in a study of differences in teaching styles for teachers in rural and urban schools. Sometimes random assignment is impossible because the independent variable simply cannot be placed under the control

of the researcher. (As for example in a study of the effects of an economic recession on the quality of decisions made by school boards.) Indeed it is often true that the random assignment of subjects to groups produces such an artificial situation that it possesses no ecological validity in the Bracht and Glass (1969) sense. Occasionally NTE research occurs as a result of poor planning or ineptitude on the part of the researcher. Regardless of the reasons for its use, one of the inevitable problems of NTE research is the confounding of "nuisance" variables with other variables.

Nuisance variables are variables that interfere with the relationships among the principal variables under study. In the simplest case they inject alternate possibilities for causal statements relating independent and dependent variables. Campbell and Stanley's sources of internal invalidity are examples of nuisance variables, in that they provide alternate explanations for differences that are observed. Other variables are identified as nuisance variables only within the contest of a particular experiment. For example, Anderson (1971), in a study to determine the effects of course content and teacher sex on classroom climate, treated the variables of class size, girl/boy ratio and class mean IQ, as nuisance variables. Wessman (1972) treated IQ, age, and race as nuisance variables in a study of the effectiveness of a compensatory education program. Aiken (1972) described several studies relating language factors to learning in mathematics. Among the nuisance variables that occurred were initial mathematics ability, IQ, and computational ability.

That the confounding of nusiance variables in the relationship between independent and dependent variables is regarded as an important

problem in educational research is well documented by the advice provided by reviewers and integrators of research in various areas. St. John (1970) for example stated "If school quality and family background are positively related to the achievement of minority pupils and to their schools' racial composition, it is crucial to control them in any study of the influence of ethnic composition (and school performance)." Welch (1969) after looking at several evaluation designs for mathematics curriculum studies, and finding a preponderance of uncontrolled studies, suggested that analysis of covariance is one of the many techniques that curriculum evaluators can use to improve their investigations. Kerlinger (1967) too, noted that "the necessity of controlling extraneous independent variables is particularly urgent in field experiments.."

There are three procedures that have been prescribed for problems of confounded nuisance variables in NTE research: matching, partial correlation, and analysis of covariance. Basically, they all attempt to answer the same question, if the nuisance variables were controlled, what would be the relationship between the independent and dependent variables?

Meehl (1970), considered the situation and suggested that the solutions may be generating more problems than they solve. In the present paper an attempt is made to extend Meehl's discussion, and to show the implications that it has for educational research. Some of the problems to be discussed have been covered in part of Elashoff (1969) and by Evans and Anastasio (1968). Problems associated with unreliability of measurement and violations of assumptions have been set out by Lord (1963) and Glass et al (1972), and will not be reiterated here.

Meehl's Analysis

Meehl noted that the practice of controlling nuisance variables in ex post facto experiment has such serious defects, that it is probably worthless for most scientifically interesting purposes. He lists three reasons.

1. Systematic unmatching. Suppose a study were conducted in which the incomes of high school graduates and dropouts are compared. If the graduates have higher income it could be argued that this results from differences in IQ; graduates having higher IQ's than dropouts in general. If graduates and dropouts are matched pairwise on the basis of IQ, and a difference still exists in income, it might be said that the difference is not attributable to IQ. However, if we look carefully at a dropout matched with a graduate at IQ 125, we would have to admit that these are likely very different people. The dropout likely has a lower achievement need than the graduate at this level. Considering the match at IQ 90, we would surely admit that the graduate with an IQ of that level has a very high achievement need. Meehl suggests that what we have done by matching on IQ is to make the groups unmatched on achievement need.

2. Unrepresentative subpopulations. When matched groups are formed using a nuisance variable which is highly correlated with one of the variables of interest, what we do in effect is to identify samples from subpopulations that differ from the entire population of interest. For example, in studying the differences between teaching styles of teachers in upper class, middle class, and lower class schools, it might be thought necessary to control for average IQ of students, size of class, kinds of facilities available, etc. The result would be that

schools are used which are so unrepresentative of their respective social classes, that they really amount to middle class schools in upper and lower class neighbourhoods. Does the study of such schools have much to say about the relationship between teaching style, and kind of school in general? Surely, generalizations should be confined to the unrepresentative subpopulations specified by the matching operation.

3. The causal arrow ambiguity. When correction operations such as matching and partial correlation are carried out, implicit assumptions about the causal direction between the nuisance variables and the dependent variable are commonly made. Very often in the social sciences this assumption is unwarranted. For example, in investigating the relationship between ethnic background and intelligence, we might decide to control for social class, under the assumption that social class and ethnic origin are related, and that social class in some sense determines intelligence. However, if the causal arrow connecting social class and intelligence actually goes the other way, i.e., intelligence determines social class to some extent, then by controlling for social class, we are reducing a valid relationship between ethnic origin and intelligence.

Although all of the defects are serious, the one of most immediate concern to the present paper is the first.

The Problem of Systematic Unmatching: Redefining the Independent Variable

Of the three difficulties that Meehl associates with the use of control in NTE research, the most telling, and complex is the problem with systematic unmatching. In his example, Meehl describes a situation in which matching is used, however the consequences are no different

when either partial correlations or analysis of covariance is used. In fact Bay and Hakstian (1972) have shown that in the case of two treatments, the significance test for the analysis of covariance is algebraically equivalent to the significance test for the partial point biserial correlation, i.e. the correlation between the treatment (expressed as 1,0) and the dependent variable with the control variable partialled out. With this in mind, it is easiest to look at the problem from the simple vantage point of the partial correlation.

To understand what happens when we partial the effects of a variable out of the relationship between two other variables, we must realize that the partial correlation is merely the correlation between two residuals. Suppose that X_1 and X_2 are two variables of interest. X_3 is a nuisance variable. Using simple regression techniques, we could predict X_1 from X_2 from X_3 as follows:

$$\hat{X}_1 = b_{13} X_3 + a$$

$$\hat{X}_2 = b_{23} X_3 + c$$

If $R_1 = X_1 - \hat{X}_1$, and $R_2 = X_2 - \hat{X}_2$, then the correlation between R_1 and R_2 is known as the partial correlation, and is symbolized $r_{12.3}$ and is equal to:

$$\frac{r_{12} - r_{13}\, r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{13}^2)}} \qquad (1)$$

and we are eliminating X_3 in this algebraic sense.

A second kind of correlation, that is of interest here, is the part correlation. If we had a variable X_4, and correlated it with R_1, then the result would be a part correlation. (It is the correlation

between variable 4 and that part of variable 1 which is uncorrelated with variable 3.) The expression for a part correlation is shown below.

$$r_{(1.3)4} = \frac{r_{14} - r_{13}\, r_{34}}{\sqrt{1-r_{13}^{2}}} \qquad (2)$$

It is Meehl's argument that when you calculate a partial correlation between two variables of interest, you tend to make $r_{(1.3)4}$ greater than r_{14}, i.e. you <u>increase</u> the relationship between the independent variable, and some outside variable. This can be illustrated by supplying some ficticious but plausible values to the relationships among the variables used by Meehl.

	1	2	3
1. Amount of Schooling	1.00	.70	.60
2. Income		1.00	.40
3. Intelligence			1.00

In the study, we would have noticed that amount of schooling and income have a fairly high correlation of .7, but if we partial out the effects of intelligence, the partial correlation $r_{(12.3)} = .628$. This indicates that even with intelligence partialed out, there is still a substantial correlation between amount of schooling and income.

Now suppose that we construct some plausible values for achievement need that reflect the effect of Meehl's example.

	Amount of Schooling	Income	Intelligence
Achievement Need	.40	.40	00

Notice that the correlation between the amount of schooling and achievement need is a modest .40. If we look at the part correlation between achievement need and amount of schooling with intelligence

partialed out, we find that the correlation is .50. Meehl argues that when you control for the effects of a nuisance variable, you can increase the influence of some other variable. In this example, the correlation involving achievement need changed from .4 to .5. Disregarding the insignificant (in a scientific sense) size of this change, we would argue that what happens when partial correlations are calculated, is that the variables of interest are transformed into new variables, i.e., the X's are changed into $X - \hat{X}$, and whereas the old variable X might have been relatively unrelated to some outside variable the new variable R, can have quite a substantial relationship with the outside variable. In our example, the newly created variable "Amount of schooling with IQ held constant" has a higher correlation with achievement need than did the old variable "Amount of Schooling".

If we consider the formula for part correlations (equation 2), we can see why this should be:

Variable 1 is one of the variables of interest (perhaps the independent variable). Variable 2 is the other variable of interest.

Variable 3 is the nuisance variable or <u>control</u> variable that is identified by the investigator for control.

Variable 4 is an outside variable that is not considered in the experiment.

Now, r_{13} is likely to be a pretty healthy correlation, since it is a nuisance variable of sufficient degree to have attracted attention. The expression $\sqrt{1 - r_{13}^2}$, will be less than one, and the larger r_{13} is, the smaller $\sqrt{1 - r_{13}^2}$ will be. Thus the denominator of the expression for the part correlation will have the effect of "magnifying" the numerator. In Meehl's example, the value of r_{34} was zero, and the effect of partialing out intelligence was to increase the correlation between

the independent variable and the outside variable from .4 to .5. It might be argued, that any outside variable that is correlated as much as .4 with the independent variable, would itself be identified as a nuisance variable and steps could be taken to control for it. On the other hand, if the correlation between the outside variable and the control variable is near zero, and the correlation between the independent variable and the outside variable is near zero, then of course there is no problem, because the numerator of the expression is zero, and no amount of magnification by $\sqrt{1 - r_{13}^2}$ can change that.

The more serious problem arises when r_{14} is very near zero, and r_{34} is magnified by

$$\frac{r_{13}}{\sqrt{1-r_{13}^2}}$$

When r_{14} is near zero, it will almost surely be overlooked in the theory that binds variables 1, 2 and 3. Of course, if r_{34} is very large, then variable 4 will be dragged to our attention on the coattails of variable 3. The area in question occurs when r_{14} lies in the interval -.2 + .2, and r_{34} is fairly small, say less than .4 in absolute value.

Consider the following hypothetical example of a curriculum evaluation situation in which there are two competing curricula. The correlations are

Curriculum (A and B)	1.00			
Achievement	.80	1.00		
Intelligence	.50	.60	1.00	
Amount of Teacher Assistance	.20	.20	-.40	1.00

The correlations with the treatment variable are point biserials. There is an apparent difference between the groups on intelligence, so an analysis of covariance is suggested. This is the same as calculating the partial point biserial correlation between treatment and achievement with intelligence partialed out. After plugging the values into the formula, $r_{12.3} = .72$, so it is concluded that even with intelligence partialled out, there is a treatment effect. But if we look at what happens to the relationship with amount of teacher assistance, we find that strange things have happened. In most classrooms it is probably true that there is a small but negative correlation between amount of assistance given and intelligence. Less intelligent students need and get more teacher assistance than more intelligent students. Let us suppose that in this case, the teacher in the treatment which had higher achievement tended to give more help than the other teacher. The part correlation between treatment with intelligence partialled out and amount of teacher assistance turns out to be .46. By partialling out intelligence, we have managed to increase the relationship between treatment and amount of teacher assistance from .2 to .46.

In a sense what we have done is to redefine the treatment variable to be "Treatment with Intelligence Partialled Out", and instead of clarifying the relationship between treatment and achievement, we have confused the issue because we have increased the confounding influence of "Amount of Teacher Assistence".

It might be suggested that the hypothetical example is "fixed" to paint the picture as darkly as possible, and this is true. Let us try to determine the conditions under which we should be alarmed. For reasons stated earlier, the problem seems to occur when $r_{(1.3)4}$ becomes greater than .4 in absolute value, and when r_{14} is less than .2 in

absolute value, and when r_{34} is less than .4 in absolute value. These conditions insure that the outside variable will not be considered an appropriate candidate for control, and it will not be pulled in on the coattails of the covariate. If we plot $r_{(1.3)4}$ against r_{34} for various values of r_{13}, and r_{14}, then we can begin to see where the danger lies. In Figure 1, the shaded area is the area of danger. The left hand figure shows various plots for $r_{14} = 0$, that is, where the independent variable and the outside variable have no correlation. Here we can see that problems could occur when the correlation between the covariate and the independent variable become .8 or greater in absolute value. In educational research these kinds of correlations don't happen very much. For curriculum evaluation, it seems inconceivable that we would select two groups for comparison if they differed by so much on a nuisance variable.

- - - - - - - - - - - - - -
FIGURE 1 About Here
- - - - - - - - - - - - - -

On the right side various lines are plotted for $r_{14} = .2$. (Taking the x axis as an axis of symetry, the mirror image would result for $r_{14} = -.2$). In this instance, some plausible situations occur. If r_{13} is as small as .4 in absolute value, we can boost the correlations between the outside variable and the independent variable from .2 to .4, making the groups as different on the outside variable as they were on the covariate. This kind of situation is destined to grey the head of the most placid researcher.

Some Side Issues

In passing one might ask if the use of analysis of covariance to improve power in a true experiment has the effect of producing a systematic difference in randomly equivalent groups. The answer is no as can be seen in the two group case.

Let X be the treatment variable scored 1, 0, and Y be the dependent variable for example, achievement, and Z be the covariate, IQ, and W be an outside variable like motivation. Over the long run, the expected value of r_{xz} is zero because people are randomly assigned to groups, and the covariate is measured before the treatment. Thus,

$$r_{xy.z} = \frac{r_{xy}}{\sqrt{1 - r_{yz}^2}}$$

and so the power is increased. But what happens to the partial correlation $r_{xw.z}$? To begin with, because of the random assignment of people to groups, r_{xw} should be zero in the long run. That being the case, the partial correlation turns out to be

$$r_{xw.z} = \frac{0 - 0 \times r_{wz}}{1\sqrt{1 - r_{wz}^2}} = 0$$

Consequently, in the true experimental case, there is no expected mismatch on the outside variable arising from the use of a control variable.

It is also of interest to see if the problems that arise in NTE designs and the use of analysis of covariance also arises in the use of analysis of variance in the NTE design. For example, suppose there is a two curriculum study, in which groups are not randomly assigned to

treatments. If we think that intelligence has a possible effect, we might try to control it by using it as a factor in a treatment by IQ analysis of variance. We can do this in two ways. In the first way, we might dichotomize (or use any number of levels) IQ by cutting below and above 110. If IQ is a reasonable cause for the observed difference in the dependent variable, (at this rather gross level), then we will find that there are disproportionate numbers in the cells of the two way design. If we analyse the data using least squares procedures for the unequal n ANOVA, then the problems described earlier will occur, since the analysis is no different in principle than using analysis of covariance with IQ levels and IQ by treatment interaction as covariates. And since IQ and treatment are correlated, the correction could serve to increase the correlation between the "adjusted" treatment variable and some outside variable.

If on the other hand, after we make out groups on IQ, we randomly throw out subjects to make the cell sizes equal, then we make the correlation between the independent variable and the nuisance variables zero. So that $r_{(1.3)4} = r_{14}$. The same situation holds true in any two-way design in which we have equal cell sizes, and subjects not assigned at random.

Of course in most two factor NTE designs, the factors are chosen because of the possible relationship to the dependent variable, and not because of their relationship to each other.

Discussion

In the past, design experts have been able to supply pat remedies to problems that are raised. Complexity of analytical methods has

usually outstriped advances in design methodology, and so new procedures could be prescribed to overcome the inadequacies of the old ones. In the present circumstances, however, the problems are not so easily solved. According to Cronbach and Meehl (1955) a construct derives its meaning from the network of relationships that connect it with other constructs and with observables. When we partial out, or control the effects of nuisance variables we change the nomological network in which the variables of interest are embedded. Or, put another way, the relationships among residual variables are simply not the same as the relationships among the original variables. We cannot interpret the residual variables as if they had the same construct status as the original variables.

As Meehl notes, when we try to investigate the relationship between naturally occurring characteristics by controlling for some nuisance variable, we are almost inevitably lead to the assertion of a counterfactual conditional statement as: If dropouts and graduates had the smae IQ, they would earn different salaries. Traditionally, researchers have acted as though the counterfactual premise (if dropouts and graduates had the same IQ) could be true in isolation from all other variables. We have seen that this is not true. When we correct for IQ we change the outside relationships. Additionally, Meehl argues that the implicit assumption does not make common sense. A society in which dropouts and graduates have the same IQ may have such different sociological mechanisms operating that the dynamics underlying income, IQ and schooling would be radically different.

Perhaps the most useful way to deal with the problem of interpreting NTE research is to live with the nuisance variables by

incorporating them into our models of behavior. By doing this, we construct models to fit the world, rather than constructing worlds to fit our models. In classroom research and evaluation, if naturally occurring characteristics confound each other, it does not make sense on either scientific or pragmatic grounds to try to isolate individual effects. The nature of the domain in which we work is complex and interactive. Our models and procedures must reflect this. Too often we have tried to overcome simplistic observation with complex analyses. It makes more sense to take more observations across time in an attempt to capture the interweaving of variables. At the present stage of sophistication, reliable description of changing relationships will lead us further than touched-up cross-sectional snapshots.

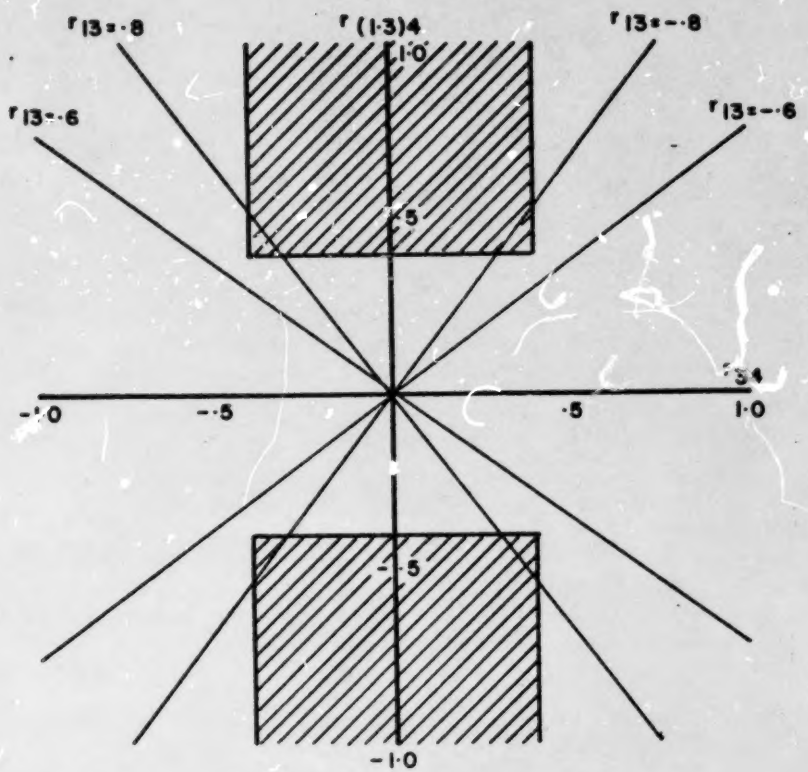

Figure 1(a): Plot of $r_{(1.3)4}$ vs. r_{34} for $r_{14} = 0$

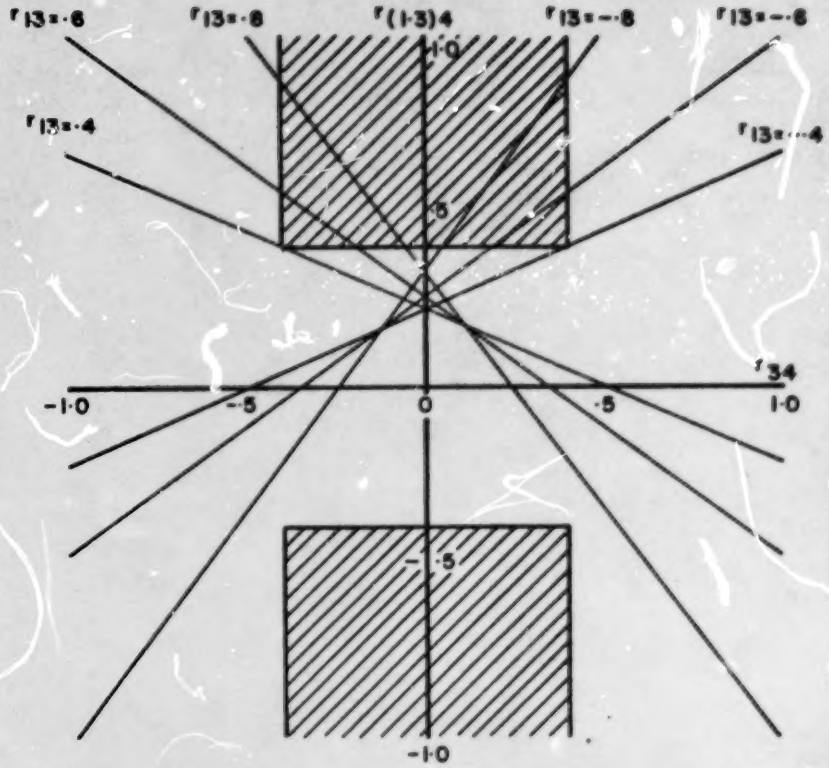

Figure 1(b): Plot of $r_{(1.3)4}$ vs. r_{34} for $r_{14} = .2$

FIGURE 1

$r_{(1.3)4}$: correlation between independent variable with covariate partialled out, and outside variable

r_{34} : correlation between covariate and outside variable

r_{13} : correlation between independent variable and covariate

r_{14} : correlation between independent variable and outside variable

REFERENCES

AIKEN, L. R. Language factors in learning mathematics. Review of Educational Research, 1972, 42, 359.

ANDERSON, G. Effects of course content and teacher sex on the social climate of learning. American Educational Research Journal, 1971, 8, 649-663.

BAY, K. S. and HAKSTIAN, A. R. Note on the equivalence of the significance test of the partial point-biserial correlation and the one-factor analysis of covariance for two treatment groups. Multivariate Behavioral Research, 1972, 3, 391-396.

BRACHT, G. H. and GLASS, G. V. The external validity of experiments. American Educational Research Journal, 1969, 5, 437-474.

CAMPBELL, D. T. and STANLEY, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally 1966 (Also appears ad Chapter 5, Experimental and quasi experimental designs for research on teaching in Gage, N. L. Handbook of Research on Teaching, Chicago: Rand McNally, 1963).

CHAPIN, F. S. Experimental designs in sociological research. New York: Harper, 1955.

CRONBACH, L. J. and Mechl, P. E. Construct validity in psychological tests. Psychological Bulletin, 1955, 52, 281-302.

ELASHOFF, J. D. Analysis of covariance a delicate instrument. American Educational Research Journal, 1969, 6, 383-401.

EVANS, S. H. and ANASTASIO, E. J. Misuse of analysis of covariance when treatment effect and covariate are confounded. Psychological Bulletin, 1968, 69, 225-234.

GLASS, G. V., PECKHAM, P. D. and SAUNDERS, J. R. Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. Review of Educational Research, 1972, 42, 237-288.

KERLINGER, F. N. Foundations of Behavioral Research. New York: Holt, Rinehart and Winston, 1967.

LORD, F. M. Large sample covariance analysis when the control variable is fallible. Journal of the American Statistical Association, 1960, 55, 307-321.

MEEHL, P. E. Nuisance variables and the ex post factro design in in Minnesota Studies in the Philosophy of Science, Volume IV, 1970, 373-402.

ST. JOHN, N. Desegregation and minority group performance. Review of Educational Research, 1970, 40, 111-134.

WELCH, W. Curriculum evaluation. Review of Educational Research, 1969, 39, 429-443.

WESSMAN, A. Scholastic and psychological effects of a compensatory education program for disadvantaged high school students: Project ABC, American Educational Research Journal, 1972, 9, 361-372.